Supplementary Material

Joanna I. Sułkowska, Piotr Sułkowski, Piotr Szymczak and Marek Cieplak

In this Supplementary Material we present more details concerning the geometry and properties of the knotted proteins 1o6d and 1v2x. Details of the simulation model and algorythm to determined the knot position during unfolding.

1 Geometry of proteins 1o6d and 1v2x

Homodimeric α/β-knot metylotransferase (Mtases) from escherichia coli YbeA (1o6d), consists of 5 α helices which surround the β sheet built of 6 β strands. The secondary structure elements are referred to as follows: 1 – β strand (amino acids 2-8); 2 – α helix (12-27). The structure is knotted between β strands 5 and 10.

2 Pulling by various p_1 and p_2 in protein 1o6d

Here we discuss several further examples of pulling by various p_1 and p_2.

2.1 Pulling by p_1 inside and p_2 outside the knotted core

Firstly we consider p_1 inside the knotted core, and p_2 outside this core, as in Fig. 2 (the second situation). More precisely, we consider p_2 equal to either 122 or 146 and p_1 equal to 73, 76, 94, and 103 for each of these choices of p_2. The corresponding simulations performed at $T = 0$ always lead to knot tightening in agreement with the geometry based prediction (this also proves the assymetry in untying from both ends: as discussed above, pulling by $p_2 = 101$ and $p_1 \leq 43$ instead leads to untying). On the other hand, for $T = 0.3\epsilon/k_B$ and p_2 close to $k_2 = 119$ we find both knotted or unknotted final conformations. Analysis of mechanical resistance and unfolding scenarios indicates that some backtracking (i.e. a long series of breaking and reforming of a group of contacts) appears during unraveling and 'decides' about the final outcome. Rupturing of the mechanical clamp

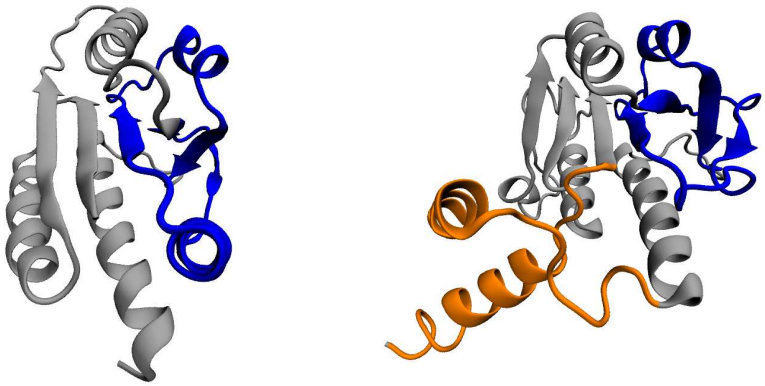

Figure 1: Proteins 1o6d (left) and 1v2x (right). For each protein a region where the knot is localized is shown in blue. The extension of the terminal regions in the case of protein 1v2x (with respect to those in 1od6) is shown in orange. The characteristic accessible points in protein 1v2x are: 20, 31, 48, 59, 67, 80, 93, 103, 108, 112, 125, 138 145, 179, 191.

around the force peak does not in itself lead to unknoting but it generates a big loop between amino acids 46 and 76, if p_2 =76, or between 63 and 94 if p_1 =94. The knot gets untied when the N-terminus retracts through one of these loops.

2.2 Pulling by p_1 and p_2 inside the knotted core

As another situation, we discuss pulling by p_1 and p_2 inside the knotted core, as shown in Fig. 2 (the third situation) in the main text. We consider $p_1 = 72$ and $p_2 = 114$, as well as a few other situations with larger p_1 and smaller p_2. As we first show, pulling by $p_1 = 72$ and $p_2 = 114$ should lead to the knotted tightened conformation, so for all those other situations the result would be the same.

In the initial configuration, the size of the knotting loop is

$$k_2 - k_1 = 119 - 65 = 54.$$

First of all, during stretching by $p_1 = 72$ and $p_2 = 114$ this loop will enlarge, which is argued as follows. When we stretch the protein to the maximal distance, the part of the chain between p_1 and p_2 ultimately becomes a straight

interval of length $114 - 72 = 42$. Then the knotting loop turns to a degenerate triangle, whose one side is this interval between p_1 and p_2, and two other sides are parallel to it and their sum is also equal to $p_2 - p_1$ (they meet in the point where the knotting occurs). Therefore the size of this degenerate loop is

$$2(p_2 - p_1) = 2 * 42 = 84.$$

This means that the knotting loop is enlarged by $84 - 54 = 30$ amino acids (as long as it persists).

Because of this enlargment of the knotting loop, the lengths of the two free ends of the knot get shortened correspondingly. In the native configuration these lengths, in the sequential units, are equal to $k_1(t) - 1$ and $N - k_2(t)$ respectively, where $k_i(t)$ denotes the time-dependent positions of the endpoints of the knot. If there are few contacts connected to the free ends (or if they are missing like in a homopolymer), one can assume that the length shortening proceeds at an approximately same rate at each end. Now note that the attachment points $p_1 = 72$ and $p_2 = 114$ are nearly in the same sequential distance ($\simeq 6$) to the closest ends of the knot ($k_1 = 65$ and $k_2 = 119$) in the native state. Thus if the knotting persists, it is likely that in the final stretched situation, the ends of the expanded knot should be in a similar distance from the attachment points. Since $p_2 - p_1 = 42$, this knotting should be around $42/2 = 21$ amino acids from p_1 and p_2. Therefore the knot ends should be found at positions $p_1 - 21 = 51$ and $p_2 + 21 = 135$. Both of these points still belong to the protein chain (which has a length of 147), so the knotting will persist. This is indeed observed by our simulations at $T = 0$ as described in more detail below, and therefore confirms correctness of the assumption that the lengths of the free ends get shortened at similar rates.

We also note that the above assumption is, to some extent, also a consequence of the fact that the jumps of the knot ends during stretching are not very abrupt (and which in principle can be abrupt, as observed e.g. for protein 1j85, see ref. [8] in the main text), as seen from the trajectories of knot ends in Fig. 2 and 3 (second panels) which are smooth.

We also discuss in slightly more detail the outcome of our simulations for various choices of p_1 and p_2. Our choices of pulling directions involved p_1=72, 76 and p_2=93, 102, 114. We have indeed found that untying is not possible at $T = 0$ in agreement with the geometric predictions, while for $T = 0.3\epsilon/k_B$ we encounter various possibilities. Pulling in all directions with $p_2 = 93$ cannot untie the knot. For p_1=73,76 and $p_2 = 114$ untying is possible, because a sufficiently large loop can be formed through which the C-terminus can be dragged. For the pulling direction corresponding to 72-100,

alternative unfolding routes lead to two different final conformations, trivial or non-trivial. The unfolding pathways in this case arise from a subtle balance between the geometry of the secondary structure and two topological barriers in the direction of pulling. For the main pathway, the final conformation is knotted, even though stretching may create a large loop through which the N-terminus could thread, which would then lead to the untying of the knot. This does not happen due to the presence of native contacts between helices 2 and 11 – the left end of the knotted structure moves in the direction of the N-terminus and kp stays in the same place. Unknotting appears (in a minor pathway) when contacts between 2α-11α and 1β-4α break. This breaking allows k_2 to move into the direction of the C-terminus which leads to untying.

2.3 Pulling by p_1 and p_2 on the left side of the knotted core

Finally we consider both p_1 and p_2 to be located on the left hand side of the knot. In agreement with expectations, the structure remains knotted in simulations at such pulling directions at $T = 0$. For $T = 0.3\epsilon/k_B$ and directions corresponding to the vicinity of 3-42, some final states are surprisingly found untied. We note that in order to untie the protein in this direction, the knotted part of the protein has to be temporarily deformed due to large number of hydrogen bonds localized between the 1β and 8β strands. In around one third of all cases, the contacts between 1β-4α, 2α-11α,1β-11α strands break just after after the force reaches F_{max}. This allows to make a large open knotted loop, through which the C-terminus unfolds back, before the unstretched part of the protein is detached from the stretched part (contacts between 1-8). Then, the structure may fold back to a partially native conformation, however the knot will be absent. The difference in the unfolding pathways is not seen from the $F(d)$ plots. However, the fraction of the native contacts, Q, and radius of gyration, R_g, show different behaviors during unfolding, e.g. the average Q decreases faster for the unknotted proteins.

3 Characteristics of stretching in various directions for protein 1o6d

In Fig. 2, Fig. 3 and Fig. 4 (bottom panel) we present: the force-displacement curves, evolution of the knot endpoints, fraction of the native contacts and the radius of gyration for the 73-122, 76-122 and 63-101 pulling directions in protein 1o6d. Black and green curves represent respectively

knotted and untying pathways. In the top panel Fig. 4 we show tightened and untied conformations arises from pulling in 63-101 directions.

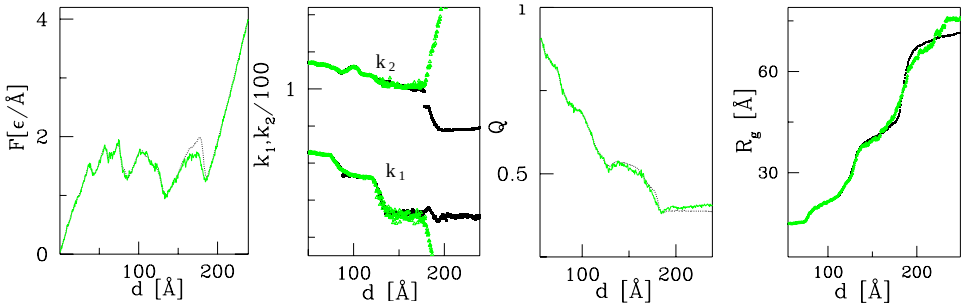

Figure 2: Stretching characteristics for the 73-122 pulling direction in 1o6d.

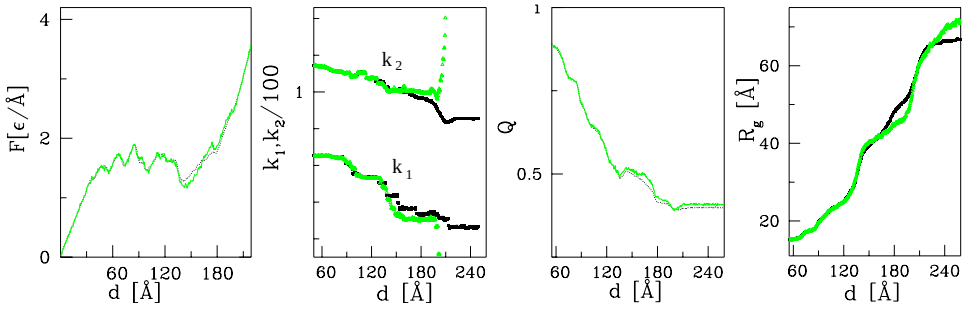

Figure 3: Stretching characteristics for the 76-122 pulling direction in 1o6d.

4 Distribution of locations of the knot endpoints for 1o6d

In the main text in Fig. 4 (right) the distribution of final locations of the knot in 1o6d is shown at $T = 0.3\,\epsilon k_B$. In Fig. 5 here, we present the same distribution at $T = 0$ and $T = 0.7\,\epsilon/k_B$.

In particular we note, that for temperatures higher than $0.3\epsilon/k_B$ we observe several changes. Firstly, there are more possible final knot configurations; some of these configurations represent a knot blocked in new, intermediate configurations (neither native nor maximally tightened). Secondly, a choice of pulling direction does not specify uniquely the final conformation. Thirdly, there are many more pulling directions for which the knot may untie.

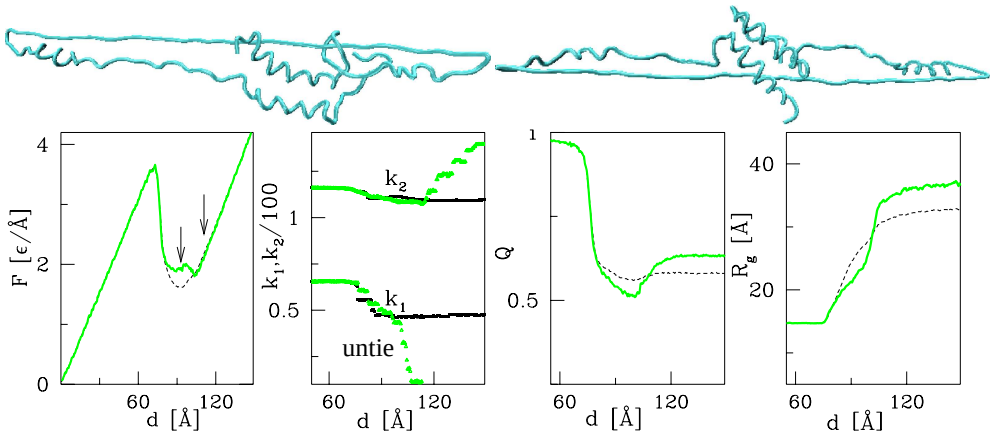

Figure 4: Top: knotted (left) or unknotted (right) final conformations upon stretching for $p_1 = 63$ and $p_2 = 101$. Bottom: stretching characteristics for the 63-101 pulling direction in 1o6d. Arrows in the first panel indicate the interval Δd in which untying occurs.

5 Pulling maps for protein 1v2x

In Fig. 6, we the present the pulling map and knotting locations for 1v2x at $T = 0$. Above the diagonal, each point has coordinates (p_1, p_2) which represent a pulling direction. The red and blue points lead, respectively, to untying or tightening of the protein, upon pulling in this direction. The red region represents an analogous deterministic region as the red wedge in the left panel in Fig. 4 in the main text. This region would spread out at higher temperatures, when stochastic character would become more and more dominant.

Below the diagonal, a map of locations of surviving knotted conformations is presented (also at $T = 0$). Here, each point has coordinates (k_2, k_1), which are sequential coordinates of a knot in the final configuration.

6 The analytical solution

The equation (2) in the main text arises from neglecting the time dependent force $F(t)$ in a more general equation

$$\frac{dP}{dt} = -k(t)P, \qquad k(t) = \frac{1}{\tau_0} e^{-\frac{E_0}{k_B T}} e^{\frac{F(t)x}{k_B T}}, \qquad t_1 < t < t_2. \tag{1}$$

As explained in the main text, E_0 is the height of the barrier at the moment $t = t_1$ (this t_1, together with t_2, are introduced in the main text; the stochastic

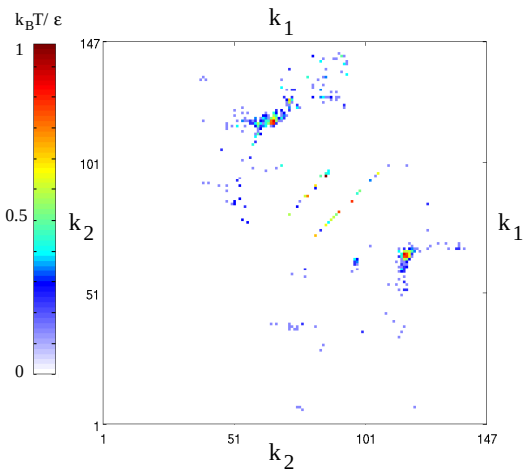

Figure 5: Position of the knot end points (k_1, k_2) in the final tightened conformations that arise from stretching at temperatures $T = 0$ (left upper part) and $T = 0.7\epsilon/k_B$ (right lower part). (The case corresponding to $T = 0.3\epsilon/k_B$ is shown in the main text). The notation is the same as in the main text.

character of untying is described by the above differential equation in the time interval between t_1 and t_2). This equation can be solved analytically by assuming that the force is linear in time (and redefined such that it vanishes at time t_1), $F(t) = cv(t - t_1)$. Such a linear behavior is observed in all of our $F(d)$ curves to hold to a good accuracy (apart from vicinity of the force maxima). This leads to the solution

$$P(t \,|\, T, v) = \exp\Big(- \frac{k_B T}{cvx\tau_0} e^{-\frac{E_0}{k_B T}} \big(e^{\frac{cvx(t-t_1)}{k_B T}} - 1\big)\Big). \tag{2}$$

The value of x cannot be determined by fitting all of our data to the full solution uniquely. We estimate it to be larger than zero and smaller than 1 Å. However, we also find that the function given in eq. (2) has an interesting property: small changes in E_0 or $c\tau_0$ can compensate for relatively large changes in x, with the entire function still being in a good agreement with the data. This means that the solution can be approximated by $x = 0$, which gives

$$P(T, v) = \exp\big(- \frac{\Delta d}{v\tau_0} e^{-E_0/k_B T}\big). \tag{3}$$

This leads to the same results as neglecting the time dependent external force, which we assumed in the main text.

We note that this approximation might arises also from the expansion

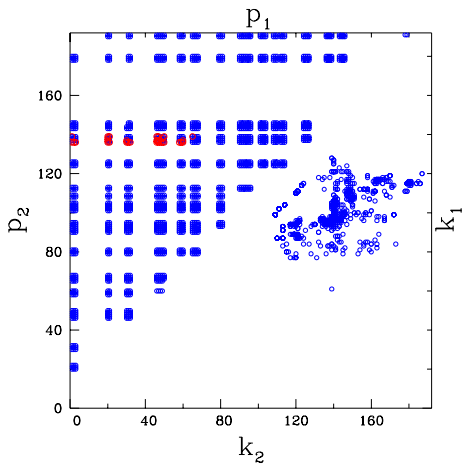

Figure 6: Pulling map and knotting locations for 1v2x at $T = 0$.

of the exponent in the last bracket ($e^{\frac{cvx(t-t_1)}{k_BT}} - 1$) and by keeping only the linear terms. In principle, this would be justified as long as the expansion parameter $w = \frac{cvx(t_2-t_1)}{k_BT}$ is sufficiently small. Let us, therefore, estimate the value of this expansion parameter. First of all, as shown in Fig. 7, knot untying (as shown by the colored intervals which indicate a change of the sequential position of $k_2(d)$ during stretching) is possible only when close to the maximum in the force, in the interval indicated by the arrow. From this figure, we estimate a change, ΔF, in the force in this time period

$$\Delta F = cv(t_2 - t_1) \sim 0.6\,\frac{\epsilon}{\text{Å}}.$$

Even though the value of x cannot be uniquely fixed, we can estimate it to be of order $x \sim 0.5$ Å. We also note that precisely such a value of the location of the barrier in the energetic landscape was determined recently in a similar situation, for a protein with a non-trivial topology, in [1]. Therefore, on substituting these values together with $k_BT \sim 0.3\epsilon$, the parameter

$$w = \frac{cvx(t_2 - t_1)}{k_BT} = \frac{\Delta F\, x}{k_BT}$$

becomes of order 1. However, this parameter gets smaller when x gets closer to 0. Thus the expansion with respect to w can be considered as a formal justification underlying the two-parameter description provided by eq. (2).

7 Structure based model

We use a simple structure based protein model with an implicit solvent. The model protein is represented by a chain of C^{α} atoms tethered along the backbone by harmonic potentials with minima at 3.8Å. Effective interactions between residues are split into native and nonnative interactions depending on whether a pair of the amino acids involved forms a native contact or not. Amino acids (i and j) that form a native contact are endowed with the effective Lennard Jones potential $V_{ij} = 4\epsilon \left[(\frac{\sigma_{ij}}{r_{ij}})^{12} - (\frac{\sigma_{ij}}{r_{ij}})^{6} \right]$ with energy scale ϵ and pair-by-pair distances r_{ij}. The length parameters σ_{ij} are chosen such that the potential minima correspond pair by pair to the native state distance between the residues. Non-native contacts are represented by hardcore repulsion to prevent entanglements. Correct chirality is imposed by the angle-dependent term in the Hamiltonian. The effective value of the energy parameter ϵ is $900K$, so the reduced temperature, $\tilde{T} = k_B T/\epsilon$, of 0.3 corresponds to the room temperature. This particular implementation is described in details in refs. [2, 3].

Thermostating is provided by the Langevin noise which also mimics random kicks by molecules of the implicit solvent. An equation of motion for each C^{α} reads $m\ddot{\mathbf{r}} = -\gamma\dot{\mathbf{r}} + \mathbf{F}_c + \mathbf{\Gamma}$, where $\mathbf{F}_c$ is the net force on an atom due to the molecular potentials and $\mathbf{\Gamma}$ is a Gaussian noise term with dispersion $\sqrt{2k_B T}$.

In our stretching simulations, the protein is attached, at amino acids p_1 and p_2, to two harmonic springs with spring constant $k = 0.06\ \epsilon/\text{Å}$. One spring is tethered, whereas the other one moves with a velocity v. Except when studying the velocity dependence, we use v=0.005 Å/ns.

In order to define the knotted core, i.e., the minimal segment of amino acids that can be identified as a knot, we use the Koniaris-Muthukumar-Taylor algorithm [4]. It involves removing the C^{α} atoms, one at a time, as long as the backbone does not intersect a triangle set by the atom under consideration and its two immediate sequential neighbors. As a result of this procedure, two end points of the knot, k_1 and k_2, are identified.

References

[1] J. Sułkowska, P. Sułkowski, J. Onuchic, *Phys. Rev. Lett.* **103** 268103 (2009).

[2] J. I. Sułkowska and M. Cieplak, *J. Phys.: Cond. Mat.* **19** 283201 (2007).

[3] J. I. Sułkowska and M. Cieplak, *Biophys. J.* **95**, 3174 (2008).

[4] K. Koniaris, M. Muthukumar, *Phys. Rev. Lett.* **66**, 2211 (1991).

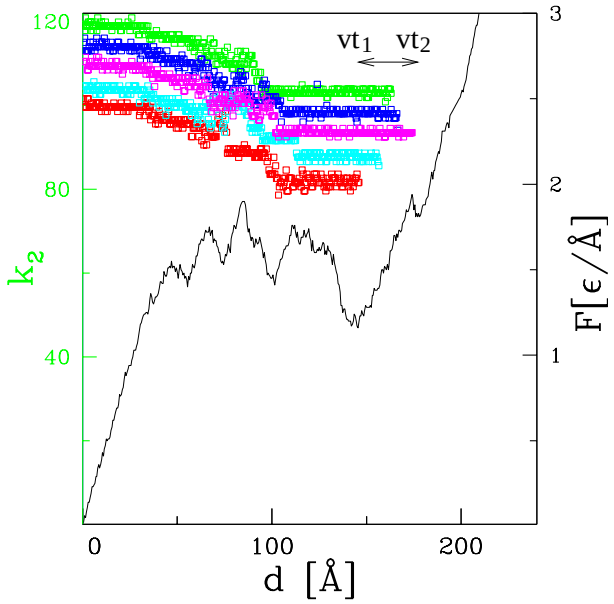

Figure 7: Stretching induced motion of the k_2 end of the knot in protein 1o6d for a direction of pulling specified by $p_1 = 76$ and $p_2 = 122$. Five sample $k_2(d)$ trajectories are indicated by the symbols in color. k_2 for the green data points is measured against the coordinates shown on the left y-axis. The other colored data points correspond to several different trajectories. They are shown displaced downward by 5, from one set to another, for clarity. At d=0, k_2 is equal to 119 in each data set. The rightmost point in each set corresponds to the instant at which the knot persists for the last time: the knot gets untied. The $F(d)$ curve is shown by the black line. This curve is averaged over and the corresponding y coordinates are indicated on the right. Note that untying occurs only on a slope before the last force peak, or in its close vicinity. The range of the pulling distances, from vt_1 to vt_2, in which untying may take place is indicated by the arrow.